DOCUMENT RESUME

ED 093 977 TM 003 833

AUTHOR	Draper, John F.
TITLE	On the Expectations of Mean Squares Based on Nonindependent Variates in Factorials.
PUB DATE	[Apr 74]
NOTE	30p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th, Chicago, Illinois, April 1974)
EDRS PRICE	MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS	*Analysis of Variance; *Expectation; *Hypothesis Testing; Mathematical Models; Matrices; Statistical Analysis

ABSTRACT

A study was made of the problem of representing the expectations of mean squares associated with analysis of variance sources of variation for experimental designs. These designs have a factorial structure over repeated measures or, for some other reason, have variates within a factorial design not all of which are mutually independent. A simple means of expressing these expectations was developed, and an example is presented from which generalization should be apparent. Implications are discussed. (Author)

ON THE EXPECTATIONS OF MEAN SQUARES BASED ON NONINDEPENDENT VARIATES IN FACTORIALS

John F. Draper

CTB/McGraw-Hill

A paper presented at the

American Educational Research Association meetings in Chicago

April 1974

On the Expectations of Mean Squares Based on Nonindependent Variates in Factorials

John F. Draper

CTB/McGraw-Hill

Several procedures have been published for aiding an investigator who wishes to perform an analysis of variance (ANOVA) on data which can be assumed to have zero pairwise covariance (i.e., Cornfield & Tukey, 1956 and Millman & Glass, 1967). Others have published methods for modifying basic ANOVA procedures so that they may be employed on data which have nonzero pairwise covariance between data elements associated with the levels of one factor in a factorial experimental design (i.e., Greenhouse & Geisser, 1959). In all of these procedures it is essential to determine the expectation of each of the mean squares associated with sources of variation in the design both under null effect and non null effect conditions. Only if these expectations are known can any appropriate variance ratio tests of effects due to a source of variation be determined. For in order to assume a central F distribution for a variance ratio test statistic which is testing a null effect, it is necessary to assume with assurance that the numerator and the denominator have identical expectations under conditions of null effects associated with the source of variation being tested.

As it will be shown with an example later in this paper, there are cases of factorial experimental designs in which a variance ratio test statistic has identical expectations for numerator and denominator under null effect conditions when based on uncorrelated variates but which has unequal expectations for numerator and denominator under null effect conditions when based on variates whose pairwise covariances are not all identically zero. That

is, a test statistic which is unbiased when based on uncorrelated variates is biased when based on correlated variates. For this reason it is essential, if an ANOVA is being considered, to be able to determine the effect of nonzero covariance among the variates on the expectation of mean squares in factorials.

It is relatively well known that the n-dimensional array of variates for an n-way factorial experimental design may be mapped into a one-dimensional array or vector, (i.e., $Y = \underline{z}$, the n-dimensional array Y is mapped into the vector $\underline{z}$) and that the ANOVA linear model may then be written in matrix form as

(1) $$\underline{z} = X\,\underline{\Upsilon} + \underline{e} \;,$$

where X is a "design matrix," $\underline{\Upsilon}$ is a vector of effect parameters, and $\underline{e}$ is a vector of error variates.

If X is a full design matrix and is written in terms of zeros and unities, it will be of deficient column rank. Therefore, unique values of the elements of $\underline{\Upsilon}$ cannot be determined. However, if appropriate contrasts are employed for the sources of variation, a reparameterized design matrix of full column rank, K, may be specified and X factored into the matrix product KL,

where

(2) $$X = KL \;,$$

and where

(3) $$L = (K'K)^{-1}K'X \;.$$

Then a new linear model for the ANOVA in terms of the contrasts may be written as

(4) $\underline{z} = KL\,\underline{\gamma} + \underline{e}$

or, say

(5) $\underline{z} = K\underline{\theta} + \underline{e}$

where $\underline{\theta} = L\underline{\gamma}$ and contains a new set of effect parameters which are linear combinations of the parameters in $\underline{\gamma}$.

The columns of K may be partitioned into sets of columns such that each partition is associated with a source of variation in the ANOVA and the sum of squares associated with that source of variation can be expressed as a quadratic form involving that partition of K and $\underline{z}$. For example

(6) $SS_A = \underline{z}' K_A (K'_A K_A)^{-1} K'_A \underline{z}$

or say

(7) $SS_A = \underline{z}' F_A \underline{z}$,

where F_A stands in for the idempotent matrix of the form.

The expectation of the mean square associated with the A source of variation, $E(MS_A)$, is equal to the expectation of the sum of squares, $E(SS_A)$ divided by the appropriate degrees of freedom,

(8) $E(MS_A) = df^{-1}\, E(SS_A)$.

The expectation for the sum of squares can be written in terms of the quadratic form as

$$(9) \qquad E(SS_A) = E \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} \; z_i \, z_j \quad ,$$

or removing the constants from the expectation

$$(10) \qquad E(SS_A) = \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} \, E(z_i z_j) \; .$$

Since the covariance of z_i and z_j , σ_{ij}, is expressed by the identities

$$(11) \qquad \sigma_{ij} = E(z_i z_j) \; - \; E(z_i) \; E(z_j)$$

and

$$(12) \qquad E(z_i z_j) = E(z_i) \; E(z_j) + \sigma_{ij},$$

The expression for the expected sum of squares may be rewritten as

$$(13) \qquad E(SS_A) = \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} \; (E(z_i) \, E(z_j) + \sigma_{ij})$$

or as

$$(14) \qquad E(SS_A) = \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} \; E(z_i) \; E(z_j) + \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} \, \sigma_{ij} \; .$$

If the covariances are all zero then the

$$(15) \qquad E(SS_A) = \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} \; E(z_i) \; E(z_j) + \sum_{i=1}^{n} f_{ii} \; \sigma_{ii} \; .$$

However, if the covariances are not all zero, the term

$$(16) \qquad 2 \sum_{i=1}^{(n-1)} \sum_{j=i+1}^{n} f_{ij} \; \sigma_{ij}$$

must be added to the expression which obtains under the usual assumptions of zero covariance.

What has been shown up to this point is that the determination of the expectation of mean squares based on variates with nonzero covariance can be accomplished by employing the same methods which apply when zero covariance is assumed, if a term involving the sum of products of the off diagonal elements of an idempotent matrix of the form and the off diagonal elements of the covariance matrix is added.

The development of this paper so far could be useful but is somewhat computationally inconvenient. Fortunately, further development results in considerable simplification, both computationally and conceptually.

It will be shown that an idempotent matrix of the form, for a given source of variation, is a patterned matrix and has elements which come from a small set of easily determined constants, and that because of the way in which it is patterned the ranges of summation under which the elements do not change are also easily determined, for most designs. Note, that for ranges of the double summation,

(17) $$\sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} \sigma_{ij} ,$$

within which f_{ij} does not change in value, the f_{ij} constant may be removed from the summation as,

(18) $$f_{ij} \sum_{i} \sum_{j} \sigma_{ij} .$$

If ξ_A is allowed to represent the expectation of the mean square, $E(MS_A)$ given no nonzero covariance among the variates on which it is based, and there are k ranges of double summation within which f_{ij} does not change,

then an expression for the expected mean square, $E(MS_A)$, given some nonzero covariance among the variates can be written as,

$$(19) \qquad E(MS_A) = \xi_A + df^{-1} \sum_{p=1}^{k} n_p f_p \left(\frac{1}{n_p} \underset{\substack{\text{range}\\ p \text{ of } i\\ \text{and } j}}{\sum_i \sum_j} \sigma_{ij} \right) ,$$

where n_p is the number of elements in the p^{th} range of summation, f_p is the constant value of f_{ij} in the p^{th} range, and $\frac{1}{n_p} \sum_i \sum_j \sigma_{ij}$ is the average covariance in the p^{th} range of i and j.

Or if the product $n_p\ f_p$ divided by the degrees of freedom is considered as a coefficient for the p^{th} average covariance term,

$$(20) \qquad c_p = df^{-1}\, n_p\, f_p ,$$

and the p^{th} average covariance is indicated by $\bar{\sigma}_p$, then the general equation for the expected mean square can be expressed as

$$(21) \qquad E(MS_A) = \xi_A + \sum_{p=1}^{k} c_p\ \bar{\sigma}_p .$$

The above expression is the basic formula toward which the preceeding development has been directed. With it and a knowledge of how F is patterned, a concise expression for the expectation of a mean square can be determined in terms of a familiar expression ξ_A, plus a weighted sum of certain average covariances.

It will be useful now to consider how to specify the unique elements of a patterned "F" idempotent matrix and where they are in it. In order to do this, first consider the nature of a partition of a reparameterized design

matrix associated with a given source of variation since an "F" matrix is a function of such partitions

$$(22) \qquad (\text{e.g.,} \quad F_A = K_A \, (K_A' \, K_A)^{-1} \, K_A') \quad .$$

For any completely crossed and balanced n-way factorial design, the "K" partition associated with any source of variation can be expressed as the Kronecker or direct product of n full column rank matrices each respectively associated with a factor of the factorial. The n full column rank matrices are ordered from left to right in the product in order of increasing "rapidity of subscript change"[1] for each factor in the n-dimensional "Y" array elements from which the $\underline{z}$ elements were mapped. For example, if the source A was in a 3-way factorial and its "subscripts changed most slowly," K_A might be equal to

$$(23) \qquad \left(A \otimes \underline{1}_q \otimes \underline{1}_p \right),$$

where $\underline{1}_q$ is a vector of q unities.

For any completely balanced design with nested factors a combination of two or more partitions obtained from a full rank design matrix, created as though the factors were crossed, can be used to structure the appropriate idempotent matrix of the form for a nested source of variation. For example, say an "F" matrix is desired for the source, A nested with B. The "F" matrices

---

[1] See Appendix A for an explanation.

for the A source of variation and the A by B interaction source of variation (assuming A and B were crossed) may be added to form the "F" matrix for the A within B source.

That is

(24) $$F_{A:B} = F_A + F_{AB} \quad ,$$

since

(25) $$SS_{A:B} = SS_A + SS_{AB}$$

or in terms of a sum of quadratic forms

(26) $$SS_{A:B} = \underline{z}' \, F_A \, \underline{z} + \underline{z}' \, F_{AB} \, \underline{z} \ ,$$

therefore

(27) $$= \underline{z}' \, (F_A + F_{AB}) \, \underline{z} \ ,$$

and

(28) $$= \underline{z}' \, F_{A:B} \, \underline{z} \ .$$

Of course unbalanced designs require partitioning of sets of rows within the column partitions and are thus more complicated but tractable, if not simple.

Now as an example consider the "F" matrix for the three-way interaction A by B by C in a 3-way completely crossed factorial.

(29) $$K_{ABC} = A \otimes B \otimes C$$

and

(30) $$F_{ABC} = (\, A \otimes B \otimes C \,) \, [(\, A \otimes B \otimes C \,)' \, (\, A \otimes B \otimes C \,)]^{-1} \, (\, A \otimes B \otimes C \,)'$$

(31) $$= (\, A \otimes B \otimes C \,) \, [A'A \otimes B'B \otimes C'C]^{-1} \quad (\, A \otimes B \otimes C \,)$$

(32) $$= (\, A \otimes B \otimes C \,) \, [(A'A)^{-1} \otimes (B'B)^{-1} \otimes (C'C)^{-1}] \, (\, A \otimes B \otimes C \,)$$

(33) $$= [A \, (A'A)^{-1} \, A'] \otimes [B \, (B'B)^{-1} \, B'] \otimes [C \, (C'C)^{-1} \, C'] \ .$$

In a similar manner any idempotent "F" matrix associated with a source of variation in a balanced n-way factorial design can be expressed as a Kronecker product of n separate idempotent matrices. Now the contrasts of the levels

of any factor may be specified so that, without any loss of generalization with respect to an overall ANOVA test of effects, there are at most two unique elements in an idempotent matrix such as

(34) $$A\,(A'A)^{-1}\,A' \,.$$

This means that an n-way balanced factorial will have patterned "F" matrices with a number of unique elements less than or equal to 2^n.

Now I will present an example from which I hope generalization will be apparent since I have not yet set up any "rules of thumb" for these procedures. Consider again a three-way factorial. Let the factors be completely crossed, balanced (n = 1) and be labeled A (for subjects) with levels i = 1,2, ..., s; B (for occasions of measurement) with levels j = 1,2, ..., r; and C (for measures) with levels k = 1,2, ..., t. Further let factors A and C provide random sources of variation and B a fixed source of variation. A linear model for the dependent variates in a three-dimensional array may be written as

(35) $$Y_{ijk} = \mu + a_i + \beta_j + c_k + d_{ij} + f_{ik} + g_{jk} + h_{ijk} + e_{ijk} \;,$$

where β_j, the effect due to the j^{th} occasion, is the only fixed effect. Or if a vector $\underline{z}$ is formed of the variates y_{ijk} such that the three dimensional array y is mapped into the one dimensional array $\underline{z}$, where $\underline{z}'$ =

(36) $$[z_1, z_2, \ldots, z_{srt}] = [y_{1,1,1}, y_{1,1,2}, \ldots, y_{1,1,t}, y_{1,2,1}, y_{1,2,2}, \ldots, y_{1,r,t}, y_{2,1,1}, \ldots, y_{s,r,t}] \;,$$

the linear model may be written in terms of a design matrix X, a vector of parameters $\underline{\gamma}$ and a vector of errors $\underline{e}$.

That is

(37) $\underline{z} = X\underline{\gamma} + \underline{e}$,

where $\underline{e}$ has elements corresponding to the e_{ijk} with subscripts ordered as they were for y_{ijk} ,

(38) $\underline{\gamma} = [\mu, \underline{a}', \underline{\beta}', \underline{c}', \underline{d}', \underline{f}', \underline{g}', \underline{h}']$, and

(39) $$X = \left[\; \underline{1}_{srt} \;\middle|\; I_s \otimes \underline{1}_r \otimes \underline{1}_t \;\middle|\; \underline{1}_s \otimes I_r \otimes \underline{1}_t \;\middle|\; \underline{1}_s \otimes \underline{1}_r \otimes I_t \;\middle|\; \rightarrow \right.$$
$$\left. I_s \otimes I_r \otimes \underline{1}_t \;\middle|\; I_s \otimes \underline{1}_r \otimes I_t \;\middle|\; \underline{1}_s \otimes I_r \otimes I_t \;\middle|\; I_s \otimes I_r \otimes I_t \;\right] .$$

where $\underline{1}_n$ is a vector of n unities, and I_n is an identity matrix of order n.

Now for a moment examine in Table 1 the sources of variation, degrees of freedom, and the expected mean squares associated with each source of variation under an assumption of independent variates as elements of $\underline{z}$, that is, assume $\underline{z} \sim N\ (\underline{\mu}, \sigma_e^2 I)$ under conditions of all null effects.

Then consider how the additions to the E(MS)s required under assumptions of nonzero covariance are determined.

In matrix notation the linear model for the reparameterized ANOVA is

(40) $\underline{z} = K\,\underline{\theta} + \underline{e}$, where

(41) $$K = \left[\; \underline{1}_{srt} \;\middle|\; U \otimes \underline{1}_r \otimes \underline{1}_t \;\middle|\; \underline{1}_s \otimes V \otimes \underline{1}_t \;\middle|\; \underline{1}_s \otimes \underline{1}_r \otimes W \;\middle|\; U \otimes V \otimes \underline{1}_t \;\middle|\; \rightarrow \right.$$
$$\left. U \otimes \underline{1}_r \otimes W \;\middle|\; \underline{1}_s \otimes V \otimes W \;\middle|\; U \otimes V \otimes X \;\right] .$$

Table 1.

SOURCES, DEGREES OF FREEDOM, AND EXPECTED MEAN SQUARES FOR THE ANOVA UNDER AN ASSUMPTION OF ZERO COVARIANCE AMONG THE DATA ELEMENTS.

Source	df	E(MS)
A	s-1	$\sigma_e^2 + r\sigma_f^2 + rt\sigma_a^2$
B	r-1	$\sigma_e^2 + \sigma_h^2 + s\sigma_g^2 + t\sigma_d^2 + st(r-1)^{-1} \sum_{j=1}^{r} (\beta_j - \bar{\beta}.)^2$
C	t-1	$\sigma_e^2 + r\sigma_f^2 + sr\sigma_c^2$
AB	(s-1)(r-1)	$\sigma_e^2 + \sigma_h^2 + t\sigma_d^2$
BC	(r-1)(t-1)	$\sigma_e^2 + \sigma_h^2 + s\sigma_g^2$
ABC	(s-1)(r-1)(t-1)	$\sigma_e^2 + \sigma_h^2$

with the contrast matrices U, V, and W each of full column rank.

Again without loss of generalization with respect to an overall ANOVA test for effects let,

$$(42)\quad U=\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ -1 & 1 & 1 & & 1 \\ 0 & -2 & 1 & & 1 \\ \cdot & 0 & -3 & & 1 \\ \cdot & \cdot & 0 & \ddots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots\ 0 & -s \end{bmatrix},\qquad (43)\quad V=\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ -1 & 1 & 1 & & 1 \\ 0 & -2 & 1 & & 1 \\ \cdot & 0 & -3 & & 1 \\ \cdot & \cdot & 0 & \ddots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots\ 0 & -r \end{bmatrix}, \text{ and}$$

$$(44)\quad W=\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ -1 & 1 & 1 & & 1 \\ 0 & -2 & 1 & & 1 \\ \cdot & 0 & -3 & & 1 \\ \cdot & \cdot & 0 & \ddots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots\ 0 & -t \end{bmatrix}$$

Now the diagonal elements of $U\,(U'U)^{-1}\,U'$ are all identically $S^{-1}\,(S-1)$ and off diagonal elements are all equal to $-S^{-1}$. For the matrix $V(V'V)^{-1}\,V'$ the diagonal and off diagonal elements are respectively $r^{-1}(r-1)$ and $-r^{-1}$ and for $W(W'W)^{-1}\,W'$ the two elements are $t^{-1}(t-1)$ and $-t^{-1}$. Where the unity vectors are involved the result obtains that all of the elements of $\underset{\sim}{1}_n\,(\underset{\sim}{1}_n{}'\underset{\sim}{1}_n)^{-1}\,\underset{\sim}{1}_n{}'$ are identically n^{-1}.

For this 3-way design any "F" matrix is a "super-super matrix" and has at most $2^3 = 8$ unique elements and thus there are eight ranges of summation within which "f_{ij}" does not change, no matter with which source of variation the "F" matrix is associated. The above-mentioned summation ranges are coherent ranges and are specified in Table 2 by super matrix notation in terms of the overall variance-covariance matrix which is itself a super-super matrix as shown in Figure 1.

Table 2.

SUMMATION RANGES IN WHICH THE "F" MATRIX ELEMENTS DO NOT CHANGE EXPRESSED IN TERMS OF THE SYMMETRIC "SUPER-SUPER MATRIX Σ" DEFINED IN FIGURE 1.

RANGE #	# OF ELEMENTS IN RANGE	SUBSCRIPT RANGE	VERBAL DESCRIPTION OF RANGE
1	srt	$\sum_{m=1}^{s} \sum_{n=1}^{r} \sum_{k=1}^{t} ((\sigma_{kk})_{nn})_{mm}$	main diagonal elements of the super-super matrix
2	srt(t-1)	$2 \sum_{m=1}^{s} \sum_{n=1}^{r} \sum_{k=1}^{t-1} \sum_{k'=k+1}^{t} ((\sigma_{kk'})_{nn})_{mm}$	off diagonal elements of the t by t main diagonal blocks
3	srt(r-1)	$2 \sum_{m=1}^{s} \sum_{n=1}^{r-1} \sum_{n'=n+1}^{r} \sum_{k=1}^{t} ((\sigma_{kk})_{nn'})_{mm}$	main diagonal elements of the t by t off diagonal blocks of the rt by rt diagonal blocks
4	srt(r-1)(t-1)	$4 \sum_{m=1}^{s} \sum_{n=1}^{r-1} \sum_{n'=n+1}^{r} \sum_{k=1}^{t-1} \sum_{k'=k+1}^{t} ((\sigma_{kk'})_{nn'})_{mm}$	off diagonal elements of the t by t of diagonal blocks of the rt by rt diagonal blocks
5	srt(s-1)	$2 \sum_{m=1}^{s-1} \sum_{m'=m+1}^{s} \sum_{n=1}^{r} \sum_{k=1}^{t} ((\sigma_{kk})_{nn})_{mm'}$	main diagonal elements of off diagonal rt by rt blocks
6	srt(s-1)(t-1)	$4 \sum_{m=1}^{s-1} \sum_{m'=m+1}^{s} \sum_{n=1}^{r} \sum_{k=1}^{t-1} \sum_{k'=k+1}^{t} ((\sigma_{kk'})_{nn})_{mm'}$	off diagonal elements of the t by t diagonal blocks of the rt by rt off diagonal blocks
7	srt(s-1)(r-1)	$4 \sum_{m=1}^{s-1} \sum_{m'=m+1}^{s} \sum_{n=1}^{r-1} \sum_{n'=n+1}^{r} \sum_{k=1}^{t} ((\sigma_{kk})_{nn'})_{mm'}$	main diagonal elements of the t by t off diagonal blocks of the rt by rt off diagonal blocks
8	srt(s-1)(r-1)(t-1)	$8 \sum_{m=1}^{s-1} \sum_{m'=m+1}^{s} \sum_{n=1}^{r=1} \sum_{n'=n+1}^{r} \sum_{k=1}^{t-1} \sum_{k'=k+1}^{t} ((\sigma_{kk'})_{nn'})_{mm'}$	off diagonal elements of the t by t off diagonal blocks of the rt by rt off diagonal blocks

In the above notation for $((\sigma_{kk'})_{nn'})_{mm'}$ the inner subscripts k and k′ index measures, the next innermost set n and n′ index occasions and the outer set of subscripts index subjects.

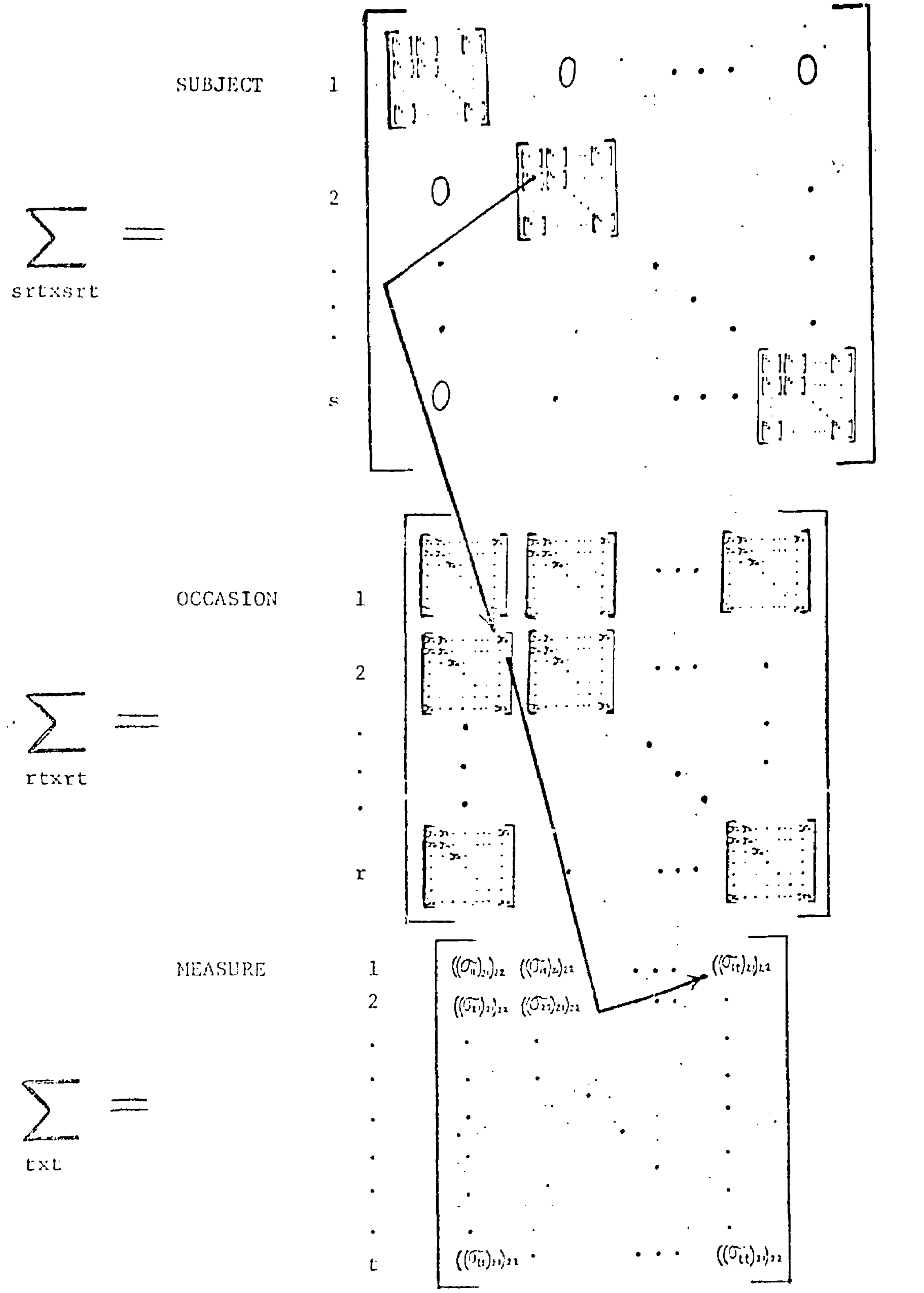

FIGURE 1

The super-super matrix representation of the complete variance-covariance matrix.

In the example shown $((\sigma_{1,t})_{2,1})_{2,2}$ is element 1,t of a matrix which is element 2,1 of a super matrix which is element 2,2 of a super-super matrix.

Table 3 contains the eight values of unique elements of the "F" matrices for each of the seven sources of variation indicated in Table 1, and Figure 1 shows a model for the variance-covariance matrix Σ which assumes zero covariance between elements involving different subjects. Let the average covariances defined by ranges 2, 3, and 4 in Table 2 be indicated respectively as $\bar{\sigma}_2$, $\bar{\sigma}_3$, and $\bar{\sigma}_4$ ($\bar{\sigma}_1$, is an average variance and $\bar{\sigma}_5$, $\bar{\sigma}_6$, $\bar{\sigma}_7$, and $\bar{\sigma}_8$ have been assumed to be identically zero). The symbols $\bar{\sigma}_2$, $\bar{\sigma}_3$, and $\bar{\sigma}_4$ represent respectively the average covariance between measures within occasions, the average covariance between the same measure on different occasions, and the average covariance between different measures on different occasions.

The values in Table 3 and expression (20) were used to create coefficients for the average covariances and the coefficients are tabulated in Table 4.

Now by employing coefficients from Table 4 and the general equation for an expected mean square which was previously developed

$$(45) \qquad \text{(e.g.} \quad E(MS_A) = \xi_A + \sum_{m=2}^{4} c_m \; \bar{\sigma}_m) \quad ,$$

for each of the sources of variation, the expected mean squares in Table 5 can be determined.

Table 3.

THE VALUES OF THE ELEMENTS OF THE "F" MATRICES FOR EACH SOURCE OF VARIATION UNDER THE EIGHT RANGES OF SUMMATION DEFINED IN TABLE 2.*

RANGE #	SOURCE						
	A	B	C	AB	AC	BC	ABC
1	$\frac{s-1}{srt}$	$\frac{r-1}{srt}$	$\frac{t-1}{srt}$	$\frac{(s-1)(r-1)}{srt}$	$\frac{(s-1)(t-1)}{srt}$	$\frac{(r-1)(t-1)}{srt}$	$\frac{(s-1)(r-1)(t-1)}{srt}$
2	$\frac{s-1}{srt}$	$\frac{r-1}{srt}$	$\frac{-1}{srt}$	$\frac{(s-1)(r-1)}{srt}$	$\frac{-(s-1)}{srt}$	$\frac{-(r-1)}{srt}$	$\frac{-(s-1)(r-1)}{srt}$
3	$\frac{s-1}{srt}$	$\frac{-1}{srt}$	$\frac{t-1}{srt}$	$\frac{-(s-1)}{srt}$	$\frac{(s-1)(t-1)}{srt}$	$\frac{-(t-1)}{srt}$	$\frac{-(s-1)(t-1)}{srt}$
4	$\frac{s-1}{srt}$	$\frac{-1}{srt}$	$\frac{-1}{srt}$	$\frac{-(s-1)}{srt}$	$\frac{-(s-1)}{srt}$	$\frac{1}{srt}$	$\frac{(s-1)}{srt}$
5	$\frac{-1}{srt}$	$\frac{r-1}{srt}$	$\frac{t-1}{srt}$	$\frac{-(r-1)}{srt}$	$\frac{-(t-1)}{srt}$	$\frac{(r-1)(t-1)}{srt}$	$\frac{-(r-1)(t-1)}{srt}$
6	$\frac{-1}{srt}$	$\frac{r-1}{srt}$	$\frac{-1}{srt}$	$\frac{-(r-1)}{srt}$	$\frac{1}{srt}$	$\frac{-(r-1)}{srt}$	$\frac{(r-1)}{srt}$
7	$\frac{-1}{srt}$	$\frac{-1}{srt}$	$\frac{r-1}{srt}$	$\frac{1}{srt}$	$\frac{-(t-1)}{srt}$	$\frac{-(t-1)}{srt}$	$\frac{(t-1)}{srt}$
8	$\frac{-1}{srt}$	$\frac{-1}{srt}$	$\frac{-1}{srt}$	$\frac{1}{srt}$	$\frac{1}{srt}$	$\frac{1}{srt}$	$\frac{-1}{srt}$

*Appendex B explains how the above values were determined.

Table 4.

COEFICIENTS OBTAINED FROM EXPRESSION 20 FOR EACH SOURCE OF VARIATION TO BE EMPLOYED IN EQUATION 21 FOR THE EXPECTED MEANS SQUARE.

RANGE #	SOURCE						
	A	B	C	AB	AC	AC	ABC
1	1	1	1	1	1	1	1
2	(t-1)	(t-1)	-1	(t-1)	-1	-1	-1
3	(r-1)	-1	(r-1)	-1	(r-1)	-1	-1
4	(r-1)(t-1)	-(t-1)	-(r-1)	-(t-1)	-(r-1)	1	1
5	-1	(s-1)	(s-1)	-1	-1	(s-1)	-1
6	-(t-1)	(s-1)(t-1)	-(s-1)	-(t-1)	1	-(s-1)	1
7	-(r-1)	-(r-1)	(s-1)(r-1)	1	-(r-1)	-(s-1)	1
8	-(r-1)(t-1)	-(r-1)(t-1)	-(s-1)(r-1)	(t-1)	(r-1)	(s-1)	-1

Table 5.

SOURCES, DEGREES OF FREEDOM, AND EXPECTED MEAN SQUARES FOR THE ANOVA UNDER THE ASSUMPTION OF NON-ZERO COVARIANCE AMONG DATA ELEMENTS OBTAINED FROM THE SAME SUBJECT.

Source	df	E(MS)
A	s-1	$\sigma_e^2 + r\sigma_f^2 + rt\sigma_a^2 + (t-1)\bar{\sigma}_2 + (r-1)\bar{\sigma}_3 + (r-1)(t-1)\bar{\sigma}_4$
B	r-1	$\sigma_e^2 + \sigma_n^2 + s\sigma_g^2 + t\sigma_d^2 + st(r-1)^{-1}\sum_{j=1}^{r}(\beta_j-\bar{\beta}.)^2 - \bar{\sigma}_3 + (t-1)(\bar{\sigma}_2 - \bar{\sigma}_4)$
C	t-1	$\sigma_e^2 + r\sigma_f^2 + sr\sigma_c^2 - \bar{\sigma}_2 + (r-1)(\bar{\sigma}_3 - \bar{\sigma}_4)$
AB	(s-1)(r-1)	$\sigma_e^2 + \sigma_h^2 + t\sigma_d^2 - \bar{\sigma}_3 + (t-1)(\bar{\sigma}_2 - \bar{\sigma}_4)$
BC	(r-1)(t-1)	$\sigma_e^2 + \sigma_h^2 + s\sigma_g^2 - \bar{\sigma}_3 + (-1)(\bar{\sigma}_2 - \bar{\sigma}_4)$
ABC	(s-1)(r-1)(t-1)	$\sigma_e^2 + \sigma_h^2 - \bar{\sigma}_3 + (-1)(\bar{\sigma}_2 - \bar{\sigma}_4)$

Examine Table 1 and Table 5 and consider the formation of a variance ratio test statistic to be employed in testing the effects due to occasions. Using either table it appears that a Quasi-F ratio (Satterthwaite, 1941) can be formed as

(46) $$\frac{MS_B + MS_{ABC}}{MS_{AB} + MS_{BC}} .$$

The Quasi-F ratio will at least have equal expectations for numerator and denominator even if an exact distribution for the statistic may not be known at present.

Now examine Table 1 and Table 5 and consider testing the subject by occasions interaction source of variation. While Table 1 provides the straightforward ratio

(47) $$\frac{MS_{AB}}{MS_{ABC}}$$

as a candidate, an examination of Table 5 indicates the expectation of the numerator is

(48) $$\sigma_e^2 + \sigma_h^2 + t\ \sigma_d^2 - \bar{\sigma}_3 + (t-1)\ (\bar{\sigma}_2 - \bar{\sigma}_4)$$

and the expectation for the denominator is

(49) $$\sigma_e^2 + \sigma_n^2 - \bar{\sigma}_3 + (-1)\ (\bar{\sigma}_2 - \bar{\sigma}_4) .$$

Thus the two expectations differ by

(50) $$t\,\sigma_d^2 + t\,(\bar{\sigma}_2 - \bar{\sigma}_4)\ .$$

Which means that if σ_d^2, the source of variation being tested, is zero, the numerator would still have an expectation which exceeded the denominator by $t\,(\bar{\sigma}_2 - \bar{\sigma}_4)$.

Therefore the ratio

(51) $$\frac{MS_{AB}}{MS_{ABC}}$$

could be a biased test statistic unless $\bar{\sigma}_2 = \bar{\sigma}_4$ which would cause the term

(52) $$t\,(\bar{\sigma}_2 - \bar{\sigma}_4)$$

to vanish, but that could only result if the average covariance between measures on the same occasion was equal to the average covariance between different measures on different occasions, which is unlikely.

Implications

The obvious implication of the developments in this paper is that an investigator who collects complicated repeated measures type data should be very careful when it comes time to decide what type of analysis is to be employed. There is a wide range of techniques from which to choose, from

a simple graphical display to complicated multivariate numerical iteration approaches, some of which provide maximum likelihood values along with the descriptive reorganization the others provide.

Univariate ANOVA procedures however do have considerable appeal. The numerical procedures involved in an ANOVA and its base of inference are widely understood. Almost anyone can do one, and a large number of investigators will not badly misinterpret one. Fortunately the bias terms such as

$$(53) \qquad t\ (\bar{\sigma}_2 - \bar{\sigma}_4)$$

can be estimated with assurance if the number of data elements is large and thus the mean squares can be "corrected" so as to form a ratio like

$$(54) \qquad \frac{MS_{AB} - t\ (\hat{\bar{\sigma}}_2 - \hat{\bar{\sigma}}_4)}{MS_{ABC}}$$

which would be an unbiased test statistic. And although I can't give the exact distribution of the above ratio when $\sigma_d^2 = 0$ at present, I can report that my simulations indicate that the Box "conservative test" for the ratio (Box, 1954) appears to be indeed conservative.

This means that a conservative investigator could use the relatively simple univariate ANOVA techniques and the developments of this paper to test hypotheses and make inferences from complicated repeated measures type data.

References

Box, G.E.P. Some theorems on quadratic forms applied in the study of analysis of variance problems: I. Effect of inequality of variance in the one way classification. *Annals of Mathematical Statistics*, 1954, 25, 290-302.

Box, G.E.P. Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effects of inequality of variance and of correlation between errors in the two-way classification. *Annals of Mathematical Statistics*, 1954, 25, 484-498.

Cornfield, J. & Tukey, J. W. Average values of mea. squares in factorials. *Annals of Mathematical Statistics*, 1956, 27, 9C -949.

Greenhouse, S.W. & Geisser S. On methods in the analysis of profile data. *Psychometrica*, 1959, 24, 95-112.

Millman, J. & Glass, G. V. Rules of thumb for writing the ANOVA table. *Journal of Educational Measurement*, 1967, 4, 41-51.

Satterwaite, F. E. Synthesis of variance. *Psychometrica*, 1941, 6, 309-16.

Rapidity of Subscript Change

If the elements of an array of data for an n-way factorial design are subscripted with n subscripts, one for each factor in the design, where each subscript indicates the level of the factor with which it is associated, a coherent mapping of the elements of the array with n subscripts into an array of elements with one subscript requires a rule for subscript correspondences. The usual rule can be shown with an example if a five-way factorial is considered with five-dimensional data array elements $y_{i,j,k,m,n}$, which are to be mapped into a one-dimensional array with data elements z_p. If lower bound values of subscripts i,j,k,m, and n are uniformly unity and the upper bound values respectively are specified as a,b,c,d, and e, then the correspondences for a balanced design can be shown by the formula

$$p = (i-1)\ bcde + (j-1)\ cde + (k-1)\ de + (m-1)\ e + n$$

or by the example

$$\begin{bmatrix} y_{1,1,1,1,1} \\ y_{1,1,1,1,2} \\ \vdots \\ y_{1,1,1,1,e} \\ y_{1,1,1,2,1} \\ y_{1,1,1,2,2} \\ \vdots \\ y_{1,1,1,2,e} \\ \vdots \\ y_{1,1,1,d,e} \\ \vdots \\ y_{1,1,c,d,e} \\ \vdots \\ y_{a,b,c,d,e} \end{bmatrix} = \begin{bmatrix} z_{1} \\ z_{2} \\ \vdots \\ z_{e} \\ z_{e+1} \\ z_{e+2} \\ \vdots \\ z_{2\cdot e} \\ \vdots \\ z_{d\cdot e} \\ \vdots \\ z_{c\cdot d\cdot e} \\ \vdots \\ z_{a\cdot b\cdot c\cdot d\cdot e} \end{bmatrix}$$

As can be seen from the example, the rightmost subscript of y changes most rapidly as the subscript for z changes, and the next rightmost subscript of y changes next most rapidly and so on.

Determination of F Elements

The values in Table 3 are easily determined once the pair of unique elements for each idempotent matrix in the Kronecker product expression for an "F" matrix analogous to expression (33) are known.

$$(33) \qquad F_{ABC} = [A(A' A)^{-1} A'] \otimes [B(B' B)^{-1} B'] \otimes [C(C' C)^{-1} C'] \quad .$$

Let the main diagonal elements of an idempotent matrix in (33) be subscripted with a unity and let the off diagonal elements be subscripted with a 2, then the unique elements of the "F" matrix in the eight ranges are determined as shown below,

Range	F Element
1	$a_1 \cdot b_1 \cdot c_1$
2	$a_1 \cdot b_1 \cdot c_2$
3	$a_1 \cdot b_2 \cdot c_1$
4	$a_1 \cdot b_2 \cdot c_2$
5	$a_2 \cdot b_1 \cdot c_1$
6	$a_2 \cdot b_1 \cdot c_2$
7	$a_2 \cdot b_2 \cdot c_1$
8	$a_2 \cdot b_2 \cdot c_2$

from which generalization can be seen.

Note also in Table 3 that when the "F" elements for the main effects have been determined, the numerators of the elements for interaction "F" matrices can be formed from the products of the numerators of the "involved" main effect "F" elements.